INTERNATIONALE HOCHSCHULE
FERNSTUDIUM

Advanced Statistics Workbook

Author: Ajay Chandra A S

Registration number: 321149868

Parameters generated on Nov 29, 2022
with signature 69b0123e50f272ec4b6304cec8ab17430a5f62c2

Date: 31-01-2023

Place: Bengaluru, India

Table of Contents

Image directory

Task 1: Basic Probabilities and Visualizations (1)

The number of meteorites falling on an ocean in a given year can be modelled by the Poisson distribution with an expectation of $\lambda = 46$. Give a graphic showing the probability of one, two, three… meteorites falling (until the probability remains provably less than 0.5% for any bigger number of meteorites). Calculate the expectation and median and show them graphically on this graphic:

The number of meteorites falling on an ocean in a given year can be modelled by a Poisson distribution because it is a discrete distribution that models the number of events in a fixed interval of time or space, given the average number of events in that interval i.e., $\lambda = 46$. The expectation of a Poisson distribution is equal to its parameter, λ ("Poisson Distribution," 2023).

As per equation 3.10 of the coursebook, the Poisson distribution $X \sim P(\lambda)$ is defined as:

$$P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!} \quad \text{where } k \in \mathbb{Z}^+ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (1.1)$$

For $\lambda = 46$, $X \sim P(46)$ will be:

$$P(X = k) = \frac{46^k e^{-46}}{k!} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (1.2)$$

To obtain the probability of one, two, three… meteorites falling, we have to calculate $P(X = 1), P(X = 2), P(X = 3)$ until the probability remains less than 0.5%. The probabilities can be calculated by substituting $k = 1,2,3 \ldots$ in the equation *(1.2)*:

$$P(X = 1) = \frac{46^1 e^{-46}}{1!} = 4.8440 \times 10^{-19}$$

$$P(X = 2) = \frac{46^2 e^{-46}}{2!} = 1.1141 \times 10^{-17}$$

$$P(X = 3) = \frac{46^3 e^{-46}}{3!} = 1.7083 \times 10^{-16}$$

.

.

$$P(X = 61) = \frac{46^{61} e^{-46}}{61!} = 0.0055 \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (1.4)$$

$$P(X = 62) = \frac{46^{62} e^{-46}}{62!} = 0.0041 \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (1.5)$$

Therefore, from equations *(1.4)* and *(1.5)* we can conclude that the probability reduces to less than 0.5% at $P(X = 62)$.

Here, the expectation is $\lambda = 46$, which means that on average, 46 meteorites are expected to fall on the ocean in a given year.

The median of a distribution of one random variable X of the discrete or continuous type is a value of k such that $P(X < k) \leq \frac{1}{2}$ and $P(X \leq k) \geq \frac{1}{2}$. If there is only one such x, it is called the median of the distribution (Hogg et al., 2005). We can calculate the median using the equation *(1.1)* as follows:

$$P(X \leq 1) = P(X = 0) + P(X = 1) \Longrightarrow 1.05306e^{-20} + 4.84408e^{-19} = 4.9494e^{-19}$$

$$P(X \leq 2) = P(X = 0) + P(X = 1) + P(X = 2) \Longrightarrow 1.1636e^{-17}$$

$$P(X \leq 3) = P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3) \Longrightarrow 1.8247e^{-16}$$

.

.

$\mathrm{P}(\mathrm{X} \leq 45) = \mathrm{P}(\mathrm{X} = 0) + \mathrm{P}(\mathrm{X} = 1) \ldots. + \mathrm{P}(\mathrm{X} = 45) \Longrightarrow 0.4804$ ………………………………………*(1.6)*

$\mathrm{P}(\mathrm{X} \leq 46) = \mathrm{P}(\mathrm{X} = 0) + \mathrm{P}(\mathrm{X} = 1) \ldots. + \mathrm{P}(\mathrm{X} = 46) \Longrightarrow 0.5391$ ………………………………………*(1.7)*

Therefore, from the equations *(1.6)* and *(1.7)*, $\mathrm{P}(\mathrm{X} \leq 45) = 0.4804$ is the closest value to 0.5 and hence $\mathrm{X} = 45$ is the required median.

To model the number of meteorites falling on an ocean in a given year using a poisson distribution, we will be defining a custom function in python on jupyter notebook. This function returns the probability mass function (PMF), which is a function that describes the probability of a random variable taking on a particular value(“Probability Mass Function,” 2022).

‘Pandas’ library is used to create the data frame, and ‘plotly_express’ is used to visualize the data.

Here is how we can use a function to plot the PMF and calculate the expectation and median of the number of meteorites falling:

```
import math
import plotly_express as px
import pandas as pd

mean = 46;
exp = 2.718281828;
x = []; y = [];
cdf_y = 0;
for i in range(0,80):
    #poisson distribution formula
    pois = ((mean**i) * exp**(-mean)) / math.factorial(i)
    x.append(i)
    y.append(pois)
# Label the axes and show the plot
df = pd.DataFrame(data=y, index=x)
fig = px.bar(df, x=x, y=y, title='Visualization of meteorites as a Poisson
distribution',
                labels=dict(x="Number of Meteorites", y="y")
            )
fig.show()
```

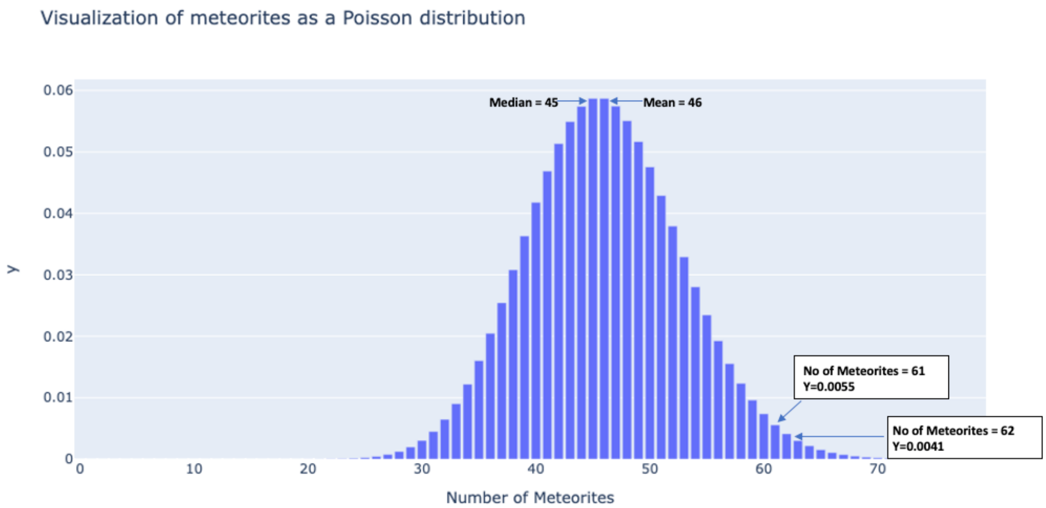

Figure 1: Visualization of meteorites in a Poisson distribution generated in jupyter-notebook

Task 2: Basic Probabilities and Visualizations (2)

Let Y be the random variable with the time to hear an owl from our room's open window (in hours). Assume that we still need to wait to hear the owl after y hours is given by the probability:

$$0.\overline{93}\,\mathrm{e}^{-3\mathrm{y}} + 0.\overline{6}\,\mathrm{e}^{-7\mathrm{y}} \ldots\ldots\ldots\ldots\ldots\ldots \text{(2.1)}$$

Find the probability that we need to wait between 2 and 4 hours to hear the owl, compute and display the probability density function graph as well as a histogram by the minute. Compute and display in the graphics the mean, variance, and quartiles of the waiting times.

The probability that we still need to wait to hear the owl after y hours is given by the equation

$$0.\overline{93}\,\mathrm{e}^{-3\mathrm{y}} + 0.\overline{6}\,\mathrm{e}^{-7\mathrm{y}} \Longrightarrow 0.9394\,\mathrm{e}^{-3\mathrm{y}} + 0.667\mathrm{e}^{-7\mathrm{y}} \ldots\ldots\ldots\ldots \text{(2.2)}$$

The probability to wait between 2 and 4 hours to hear the owl can be obtained by computing the cumulative density function (CDF) of Y over the interval 2 to 4 hours. The CDF of a real-valued random variable Y evaluated at y, is a function that maps each value of of y of the random variable to the cumulative sum of the probabilities of all values less than or equal to y ("Cumulative Distribution Function," 2023).

$$\mathrm{P}(\mathrm{a} \leq \mathrm{y} \leq \mathrm{b}) = \int_{\mathrm{a}}^{\mathrm{b}} \mathrm{f_y}(\mathrm{y})\mathrm{dy}$$

$$\mathrm{P}(\mathrm{a} \leq \mathrm{y} \leq \mathrm{b}) = \int_{2}^{4} (0.9394\mathrm{e}^{-3\mathrm{y}} + 0.667\mathrm{e}^{-7\mathrm{y}})\mathrm{dy}$$

$$= 0.9394 \int_2^4 \mathrm{e}^{-3\mathrm{y}}\mathrm{dy} + 0.667 \int_2^4 \mathrm{e}^{-7\mathrm{y}}\mathrm{dy}$$

$$= -0.3131\,\mathrm{e}^{-3\mathrm{y}}\Big|_2^4 - 0.0952\,\mathrm{e}^{-7\mathrm{y}}\Big|_2^4$$

$$= -0.3131(e^{-12} - e^{-6}) - 0.0952\,(e^{-28} - e^{-14})$$

$$= -0.3131(-2.4726\mathrm{x}10^{-3}) - 0.0952\,\,(-8.3152\mathrm{x}10^{-7})$$

$$\mathrm{P}(2 \leq \mathrm{y} \leq 4) = 0.000774 \text{ hours} \ldots\ldots\ldots\ldots \text{(2.3)}$$

Let Y be a random variable. As per definition 1.8.1 of Hogg et al., (2005), If Y is a continuous random variable with pdf $\mathrm{f(y)}$

$$\int_{-\infty}^{\infty} |\mathrm{y}|\,\mathrm{f_y}(\mathrm{y})\mathrm{dy} < \infty$$

then the expectation of Y is given by:

$$\mathrm{E[Y]} = \int_{\mathrm{a}}^{\mathrm{b}} \mathrm{y}\,\mathrm{f_y}(\mathrm{y})\mathrm{dy} \ldots\ldots\ldots\ldots \text{(2.4)}$$

$$= \int_2^4 \mathrm{y}(0.9394\mathrm{e}^{-3\mathrm{y}} + 0.667\mathrm{e}^{-7\mathrm{y}})\mathrm{dy}$$

$$= 0.9394 \int_2^4 \mathrm{ye}^{-3\mathrm{y}}\mathrm{dy} + 0.667 \int_2^4 \mathrm{ye}^{-7\mathrm{y}}\mathrm{dy}$$

$$= 0.9394 \left[\mathrm{y}\frac{\mathrm{e}^{-3\mathrm{y}}}{-3} - \frac{1}{3}\int - \mathrm{e}^{-3\mathrm{y}}\right]_2^4 + 0.667\left[-\mathrm{y}\frac{\mathrm{e}^{-7\mathrm{y}}}{7} - \frac{1}{7}\int \mathrm{e}^{-7\mathrm{y}}\right]_2^4$$

$$= 0.9394\left[\frac{-1}{3}\mathrm{ye}^{-3\mathrm{y}} - \frac{\mathrm{e}^{-3\mathrm{y}}}{9}\right]_2^4 + 0.667\left[\frac{-1}{7}\mathrm{ye}^{-7\mathrm{y}} - \frac{\mathrm{e}^{-7\mathrm{y}}}{49}\right]_2^4$$

Substituting the limits in the above equation we get,

$$= 0.9394\left[\left(\frac{-4e^{-12}}{3} - \frac{e^{-12}}{9}\right) - \left(\frac{-2e^{-6}}{3} - \frac{e^{-6}}{9}\right)\right] + 0.667\left[\left(\frac{-4e^{-28}}{7} - \frac{e^{-28}}{49}\right) - \left(\frac{-2e^{-14}}{7} - \frac{e^{-14}}{49}\right)\right]$$

$$\mathrm{E[Y]} = 0.00180 \text{ hours} \ldots\ldots\ldots\ldots \text{(2.5)}$$

The variance of a random variable Y is the expected value of the squared deviation from the mean of Y("Variance," 2022). The expression for the variance can be expanded as follows:

$$\mathrm{Var[Y]} = \mathrm{E[Y^2]} - \mathrm{E[Y]}^2 \ldots\ldots\ldots\ldots \text{(2.6)}$$

$E[Y^2]$ can be calculated using the PDF equation *(2.2)* as follows,

$$E[Y^2] = \int_a^b y^2\, f_y(y)dy$$

$$= \int_2^4 y^2(0.9394e^{-3y} + 0.667e^{-7y})dy$$

$$= 0.9394\int_2^4 y^2 e^{-3y}dy + 0.667\int_2^4 y^2 e^{-7y}dy$$

$$= 0.9394\left[y^2\frac{e^{-3y}}{-3} - \int 2y\int e^{-3y}\right]_2^4 + 0.667\left[y^2\frac{e^{-7y}}{-7} - \int 2y\int e^{-7y}\right]_2^4$$

$$= 0.9394\left[y^2\frac{e^{-3y}}{-3} - \int 2y\frac{e^{-3y}}{-3}\right]_2^4 + 0.667\left[y^2\frac{e^{-7y}}{-7} - \int 2y\frac{e^{-7y}}{-7}\right]_2^4$$

$$= 0.9394\left[y^2\frac{e^{-3y}}{-3} + \frac{2}{3}\int y\,e^{-3y}\right]_2^4 + 0.667\left[y^2\frac{e^{-7y}}{-7} + \frac{2}{7}\int y\,e^{-7y}\right]_2^4$$

$$= 0.9394\left[\frac{-1}{3}y^2e^{-3y} - \frac{2}{9}ye^{-3y} - \frac{2}{27}e^{-3y}\right]_2^4$$

$$+0.667\left[\frac{-1}{7}y^2e^{-7y} - \frac{2}{49}ye^{-7y} - \frac{2}{343}e^{-7y}\right]_2^4$$

Substituting the limits in the above equation we get,

$$= 0.9394\left[\left[\frac{-16e^{-12}}{3} - \frac{8e^{-12}}{9} - \frac{2e^{-12}}{27}\right] - \left[\frac{-4e^{-6}}{3} - \frac{4e^{-6}}{9} - \frac{2e^{-6}}{27}\right]\right]$$

$$+ 0.667\left[\left[\frac{-16e^{-28}}{7} - \frac{8e^{-28}}{49} - \frac{2e^{-28}}{343}\right] - \left[\frac{-4e^{-14}}{7} - \frac{4e^{-14}}{49} - \frac{2e^{-14}}{343}\right]\right]$$

$$E[Y^2] = 0.0042 \text{ hours}\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\textit{(2.7)}$$

Substituting the value of $E[Y^2]$ and $E[Y]^2$ in equation *(2.6)* we get,

$$= 0.0042 - (0.00180)^2$$

$$Var[Y] = 0.0041 \text{ hours}\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\textit{(2.8)}$$

In statistics, a quartile divides the number of data points into four quarters. The data must be ordered from smallest to largest to compute quartiles(“Quartile,” 2022).

The first quartile (Q_1) is defined as the middle number between the smallest number (min) and the median/middle number of the data set and 25% of the data lies below this point(“Quartile,” 2022). Since the domain for the task is taken between 2 to 4 hours, we will be integrating the PDF equation *(2.2)* from 2 to 2.5 hours (25th percentile).

$$Q_1 = \int_a^b f_y(y)\ dy$$

$$Q_1 = \int_2^{2.5}(0.9394e^{-3y} + 0.667e^{-7y})dy$$

$$= 0.9394\int_2^{2.5} e^{-3y}dy + 0.667\int_2^{2.5} e^{-7y}dy$$

$$= -\frac{0.9394}{3}e^{-3y}\ \Big|_2^{2.5} - \frac{0.667}{7}e^{-7y}\ \Big|_2^{2.5}$$

Substituting the limits in the above equation we get,

$$= -0.3131(e^{-7.5} - e^{-6}) - 0.0952(e^{-17.5} - e^{-14})$$

$$Q_1 = 0.00060 \text{ hours}\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\textit{(2.9)}$$

The second quartile (Q_2) is the median of a data set and hence, 50% of the data lies below this point(“Quartile,” 2022). Since the domain for the task is taken between 2 to 4 hours, we will be integrating the PDF equation *(2.2)* from 2 to 3 hours (50th percentile).

$$Q_2 = \int_a^b f_y(y)\ dy$$

$$= \int_2^3 (0.9394e^{-3y} + 0.667e^{-7y})dy \Rightarrow -\frac{0.9394}{3}\ e^{-3y} \Big|_2^3 - \frac{0.667}{7}\ e^{-7y} \Big|_2^3$$

Simplifying the above equation, we get:

$Q_2 = 0.00073\ \text{hours}$……………………………………………….*(2.10)*

The third quartile (Q_3) is the middle value between the median and the highest value (max) of the data set, 75% of the data lies below this point(“Quartile,” 2022). Since the domain for the task is taken between 2 to 4 hours, we will be integrating the PDF equation *(2.2)* from 2 to 3.5 hours (75th percentile).

$$Q_3 = \int_a^b f\,(y)\ dy$$

$$= \int_2^{3.5} (0.9394e^{-3y} + 0.667e^{-7y})dy \Rightarrow -\frac{0.9394}{3}\ e^{-3y} \Big|_2^{3.5} - \frac{0.667}{7} e^{-7y} \Big|_2^{3.5}$$

After simplifying the above equation, we get:

$Q_3 = 0.00076\ \text{hours}$……………………………………………….*(2.11)*

To model the probability for the waiting time to hear an owl from our room’s open window, we will be defining a custom function in python on jupyter notebook. The function returns a PDF whose value at any given point in the sample space can be interpreted as providing a relative likelihood that the value of the random variable would be equal to that sample(“Probability Density Function,” 2022).

‘Pandas’ library is used to create the data frame, and ‘plotly_express’ is used to visualize the pdf, mean and quartiles in the pdf and, the Histogram.

Here is how we can use a function to calculate and visualize the mean, quartiles, pdf and histogram of the probability of the waiting time to hear an owl from our room’s open window:

```
import plotly_express as px
import pandas as pd
list=[];
for rec in range(120,240):
    list_rec = rec/60;
    list.append(list_rec)
x = []; pdf_list = []
exp_const, f_exp_pwr, s_exp_pwr, F, G, H , I = 2.718281828, 3, 7, 0, 0, 0, 
0
for i in list:
    lwr_lim = i; upr_lim = i+1/60
    F = -abs(lwr_lim*f_exp_pwr); G = -abs(upr_lim*f_exp_pwr)
    H = -abs(lwr_lim*s_exp_pwr); I = -abs(upr_lim*s_exp_pwr)
    num1 = -abs(0.3131); num2 = -abs(0.0952)
    eqn1 = exp_const**G - exp_const**F; eqn2 = exp_const**I - exp_const**H
    eqn3 = num1*eqn1; eqn4 = num2*eqn2; pdf_eqn = eqn3 + eqn4
    x.append(lwr_lim*60)
    pdf_list.append(pdf_eqn*60)
#Label the axes and show the plot
pdf_df = pd.DataFrame(data=pdf_list, index=x)
pdf_fig = px.line(pdf_df, x=x, y=pdf_list, title='Probability Density 
Function',
                labels=dict(x="minutes", y="y")
                )
mean = 125; q1 = 147; q2 = 143; q3 = 188;
```

```
pdf_fig.add_vline(x=mean, line_width=1, line_color="green")
pdf_fig.add_vline(x=q1, line_width=1, line_color="red")
pdf_fig.add_vline(x=q2, line_width=1, line_color="red")
pdf_fig.add_vline(x=q3, line_width=1, line_color="red")
pdf_fig.show()
#show the histogram
hist_df = pd.DataFrame(data=pdf_list, index=x)
hist_fig = px.histogram(hist_df, x=pdf_list, title='Histogram')
hist_fig.show()
```

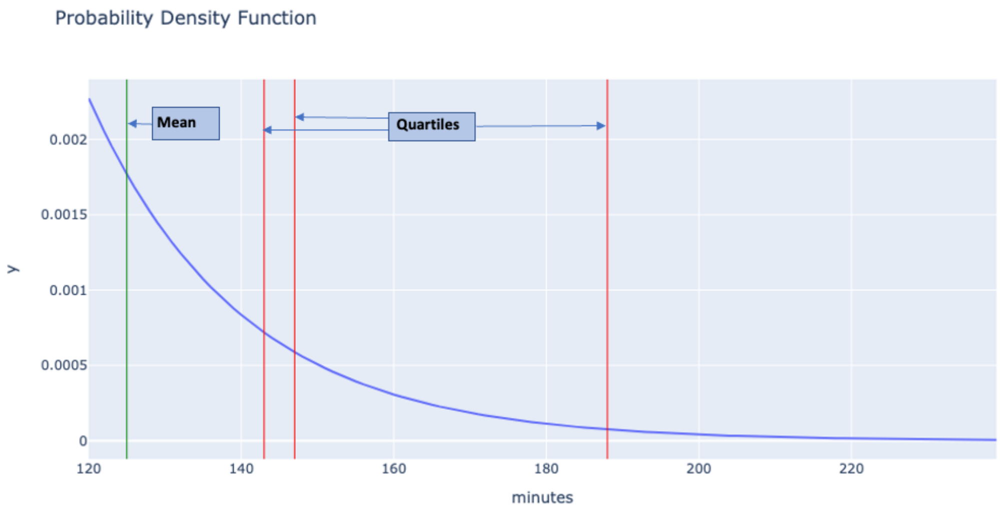

Figure 2: Probability Density Function generated in jupyter-notebook

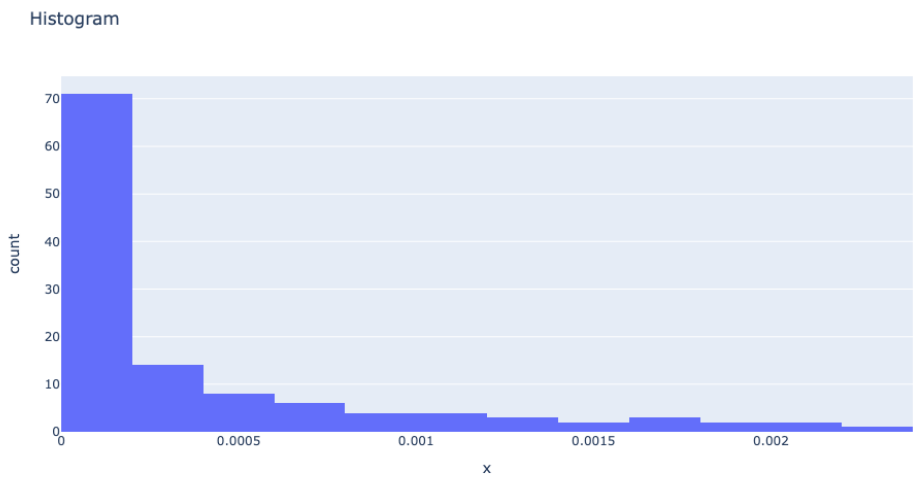

Figure 3: Histogram generated in jupyter-notebook

Task 3: Transformed Random Variables

A type of network router has a total bandwidth to first hardware failure called S expressed in terabytes. The random variable S is modelled by an exponential distribution whose density is given by the function:

$f_s(s) = \frac{1}{\theta} e^{-\frac{s}{\theta}}$ with a single parameter θ. Consider the bandwidth total to failure T of the sequence of the two routers of the same type (one being brought up automatically when the first is broken).

Express T in terms of the bandwidth total to failure of single routers S_1 and S_2. Formulate realistic assumptions about these random variables. Calculate the density function of the variable T.
Given an experiment with the dual-router-system yielding a sample $T_1, T_2, \ldots, T_n$, calculate the likelihood function for θ. Propose a transformation of this likelihood function whose maximum is the same and can be computed easily.

An actual experiment is performed, and the infrastructure team has obtained the bandwidth totals to failure given by the sequence $38, 51, 37, 4, 13$ of numbers. Estimate the model-parameter with the maximum likelihood and compute the expectation of the bandwidth total to failure of the dual-router-system.

Let the bandwidth total for the first hardware failure of a type of network router be expressed in terabytes as S. The density of the random variable S is modelled by the exponential distribution with a single parameter θ as follows:

$$f_s(s) = \frac{1}{\theta} e^{-\frac{s}{\theta}} \quad \text{……………………………………………………….(3.1)}$$

The density function for the first and second router respectively can be obtained from the given exponential distribution *(3.1)* as follows:

$$f_{S_1}(S_1) = \frac{1}{\theta} e^{-\frac{S_1}{\theta}}, f_{S_2}(S_2) = \frac{1}{\theta} e^{-\frac{S_2}{\theta}} \quad \text{…………………………………..(3.2)}$$

The bandwidth total to failure of the dual-router system, T, is the sum of the bandwidth total of the failure of the two routers, S_1 and S_2. So,

$$T = S_1 + S_2 \quad \text{………………………………………………..(3.3)}$$

It is reasonable to assume that S_1 and S_2 are independently and identically distributed, meaning that their distribution does not depend on each other and that they both have the same distribution.
Therefore, the density function of T can be obtained by the joint distribution of the first and second routers as follows:

$$f_T(T) = \frac{1}{\theta} e^{-\frac{S_1}{\theta}} \times \frac{1}{\theta} e^{-\frac{S_2}{\theta}} = \frac{1}{\theta^2} e^{-\frac{(S_1+S_2)}{\theta}} \quad \text{…………………………….(3.4)}$$

Substituting the value for T in the above equation to be the sum of S_1 and S_2 derived from equation *(3.3)*, we get,

$$f_T(T) = \frac{1}{\theta^2} e^{-\frac{T}{\theta}} \quad \text{……………………………………………(3.5)}$$

Let $T_1, T_2, \ldots, T_n$ be the samples of size ‘n’ observed from an experiment with the dual-router-system modelled using an exponential distribution $T_1, \ldots, T_n \sim \text{Exponential}(\theta)$.

As per section 6.1 of the coursebook, the likelihood is the probability of observing a given dataset or the joint distribution of the dataset evaluated at the given data. Therefore, the likelihood function for θ from the sample values $T_1, T_2, \ldots, T_n$ can be obtained by the joint distribution derived from the equation *(3.5)* as follows:

$$L[\theta] = f(T_1, \theta) \times f(T_2, \theta) \times f(T_3, \theta) \ldots \ldots \times f(T_n, \theta)$$

$$L[\theta] = \frac{1}{\theta^2} e^{-\frac{T_1}{\theta}} \times \frac{1}{\theta^2} e^{-\frac{T_2}{\theta}} \times \frac{1}{\theta^2} e^{-\frac{T_3}{\theta}} \ldots \ldots \times \frac{1}{\theta^2} e^{-\frac{T_n}{\theta}}$$

$$L[\theta] = \frac{1}{\theta^{2n}} e^{-\frac{\sum_{i=1}^{n} T_i}{\theta}} \quad \text{……………………………………….(3.6)}$$

As per section 6.1 of the coursebook, a value θ that maximizes the likelihood of observing the data is called the maximum likelihood estimate. Therefore, to obtain the maximum likelihood for the given task, we will be finding a parameter θ that maximizes $L[\theta]$ from *(3.6)* i.e., to maximize the probability of observing the given data by adjusting the value of the parameter(s) θ for an assumed choice of a probability distribution.

Since the logarithm of a product is the sum of the logarithms and it makes our likelihood function easier to handle in practice, let's take log on both sides for the equation *(3.6)* as follows:

$$\log L[\theta] = \log\left[\frac{1}{\theta^{2n}} e^{-\frac{\sum_{i=1}^{n} T_i}{\theta}}\right]$$

As per section 6.1 of the coursebook, we can compute the first derivative for the above log-likelihood function to find the maximum likelihood estimate as follows:

$$\frac{d}{d\theta}\log L[\theta] = \frac{d}{d\theta}\left[\log\left[\frac{1}{\theta^{2n}} e^{-\frac{\sum_{i=1}^{n} T_i}{\theta}}\right]\right] = 0$$

$$= \frac{d}{d\theta}\left[\log\theta^{-2n} + \log e^{-\frac{\sum_{i=1}^{n} T_i}{\theta}}\right] = 0$$

$$= \frac{d}{d\theta}\left[-2n\log\theta - \frac{\sum_{i=1}^{n} T_i}{\theta}\right] = 0$$

$$= -\frac{2n}{\theta} + \frac{\sum_{i=1}^{n} T_i}{\theta^2} = 0$$

$$\theta = \frac{\sum_{i=1}^{n} T_i}{2n} \quad \text{...}\textit{(3.7)}$$

Equation *(3.7)* is the proposed transformation of the likelihood function for θ whose maximum remains the same and can be computed easily.

The sequence of numbers for the bandwidth totals to failure obtained by the infrastructure team after performing an actual experiment are $38, 51, 37, 4, 13$.

To obtain the estimation for the model-parameter, first, we must calculate θ from the equation *(3.7)* by substituting the values for the sequence of the numbers T_i and its count 'n'.

$$\theta = \frac{\sum_{i=1}^{n} T_i}{2n}$$

$$\theta = \frac{38+51+37+4+13}{2\times 5} = 14.3 \quad \text{...}\textit{(3.8)}$$

Now, let's substitute the value of θ and T_i in the likelihood equation *(3.6)* as follows:

$$L[\theta] = \frac{1}{\theta^{2n}} e^{-\frac{\sum_{i=1}^{n} T_i}{\theta}}$$

$$L[\theta] = \frac{1}{\theta^2} e^{-\frac{T_1}{\theta}} \times \frac{1}{\theta^2} e^{-\frac{T_2}{\theta}} \times \frac{1}{\theta^2} e^{-\frac{T_3}{\theta}} \times \frac{1}{\theta^2} e^{-\frac{T_4}{\theta}} \times \frac{1}{\theta^2} e^{-\frac{T_5}{\theta}}$$

$$L[\theta] = \frac{1}{\theta^{2n}} e^{-\frac{(38+51+37+4+13)}{\theta}}$$

$$L[14.3] = \frac{1}{(14.3)^{2*5}} e^{-\frac{143}{14.3}}$$

$$L[14.3] = 1.269 \times 10^{-16} \quad \text{...}\textit{(3.9)}$$

The expected value of a random variable with a finite number of outcomes is a weighted average of all possible outcomes and is often denoted by $E[\theta]$("Expected Value," 2022a). Therefore, the expectation for the bandwidth total to the failure of the dual-router-system is given by:

$$E[\theta] = \frac{T1 + T2 + T3 + T4 + T5}{n}$$

$$E[\theta] = \frac{38+51+37+4+13}{5} = 14.3 \quad \text{..}\textit{(3.10)}$$

Task 4: Hypothesis Test

Over a long period of time, the production of 1000 high-quality hammers in a factory seems to have reached a weight with an average of 869 (in g) and a standard deviation of 26.2 (in g). Propose a model for the weight of the hammers including a probability distribution for the weight. Provide all the assumptions needed for this model to hold (even the uncertain ones). What parameters does this model have?

A random sample of newly produced hammers is evaluated yielding the weights $881, 828, 883, 863, 822, 891, 863, 919, 891, 931$ in g. Does the new system make *lower* weights?

What hypotheses can you propose to test the question? What test and decision rule can you make to estimate if the new system answers the given question? Express the decision rules as logical statements involving critical values. What error probabilities can you suggest and why? Perform the test and draw the conclusion to answer the question.

Our model for the weight of $n = 1000$ high quality hammers is a normal distribution having an average(μ) of 869 grams and a standard deviation(σ) of 26.2 grams. A normal distribution is a type of continuous probability distribution for a real-valued random variable("Normal Distribution," 2023) and the general form of its probability density function is:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(4.1)$$

The central limit theorem establishes that, in many situations, when independent random variables are summed up, their normalized sum tends towards a normal distribution("Central Limit Theorem," 2022). The weight of the hammers is a continuous variable because it can take on any value within a certain range, rather than any specific values. Additionally, since we have 1000 observations, the average of many observations of a random variable with finite mean and variance is itself a random variable whose distribution can converge to a normal distribution ("Normal Distribution," 2023). The parameters of this model are the mean and standard deviation of the weight of the hammers.

Substituting the values of μ and σ in equation *(4.1)*, we get:

$$f(x) = \frac{1}{26.2\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{x-869}{26.2})^2}$$

$$f(x) = \frac{1}{65.673} e^{-\frac{1}{2}(\frac{x-869}{26.2})^2} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(4.2)$$

X can be a standard normal distribution with $Z = \frac{x-\mu}{\sigma}$. If Z is a standard normal deviate, then $X = \sigma Z + \mu$ will have a normal distribution with expected value μ and standard deviation σ ("Normal Distribution," 2023). Since the value of n is large, we can derive the PDF equation from the equation *(4.1)* for the standardized normal variate with $\mu = 0$ and $\sigma = 1(\text{unity})$, given by:

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(4.3)$$

To test the question of whether the new system produces hammers with lower weights, we can use a left-tailed hypothesis test(The Organic Chemistry Tutor, 2019). A random sample of newly produced hammers is evaluated yielding the weights $881, 828, 883, 863, 822, 891, 863, 919, 891, 931$ in g.

Number of samples $n = 10$;

Mean of samples $\overline{X} = \frac{881+828+\cdots\ldots+931}{10} = 877.2$;

Let us define the null and the alternative hypothesis. The null hypothesis(H_0) is the claim that no difference exists between the average weight of the hammers produced from the old system and the hammers produced by the new system("Null Hypothesis," 2023). The alternative hypothesis(H_1) is the proposed proposition in the hypothesis test("Alternative Hypothesis," 2022) i.e., the average weight of the hammers produced by the new system is lower than the old system.

Null Hypothesis $H_0 : \mu = 869$ g; Alternative Hypothesis $H_1 : \mu < 869$ g

From section 7.1 of the coursebook, test statistic is given by:

$$t = \frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \textit{(4.4)}$$

where $\overline{X}$ is the sample mean, σ is the standard deviation of the population, n is the number of samples and μ is the hypothesis we want to test.

Substituting the values of $\overline{X}$, μ, σ and n in the equation *(4.4)* we get:

$$t = \frac{877.2 - 869}{\frac{26.2}{\sqrt{10}}}$$

$$t = 0.989 \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \textit{(4.5)}$$

A confidence interval (C) is a range of estimates for an unknown parameter. A confidence interval is computed at a designated confidence level; the 95% confidence level is most common("Confidence Interval," 2022). For example, out of all intervals computed at the 95% level, 95% of them should contain the parameter's true value. Since, the confidence interval is not specified in the task, we will be going with 95%.

Referring to section 7.1 in the coursebook, the error probabilities for this test are type I error probability and type II error probability. The type I error is the probability of rejecting the null hypothesis when it is true. The type II error is the probability of not rejecting the null hypothesis when it is false.

α is called the significance level, and is the probability of rejecting the null hypothesis given that it is true (a type I error) ("Statistical Significance," 2022), given by:

$$\alpha = 1 - C \Rightarrow 1 - 0.95$$

$$\alpha = 0.05 \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \textit{(4.6)}$$

From section 7.1 in the coursebook, to decide whether to reject the null hypothesis, we need to define a critical value Z_α as a cut-off. For any value of the test statistic below this value, we accept the null hypothesis, for any value above we reject the null hypothesis and accept the alternative hypothesis.

Since we have taken $\alpha = 0.05$ from the equation *(4.6),* the critical value with $10 - 1$ degrees of freedom is given by the t-table(HelpYourMath - Statistics, 2019) as follows:

$$|Z_\alpha| = 1.833 \ldots\ldots (4.7)$$

On comparing the value of the test statistic from equation *(4.5)* and the critical value from equation *(4.7)*, we see that the value of the test statistic is less than the critical value i.e., $t < Z_\alpha$.

Therefore, there is enough evidence to reject the alternative hypothesis i.e., the new system makes lower weights than the old system. Therefore, we can accept the null hypothesis H_o and reject H_1 .

Task 5: Regularized Regression

Given the values of an unknown function $f: \mathbb{R} \rightarrow \mathbb{R}$ at some selected points, we try to calculate the parameters of a model function using OLS as a distance and a ridge regularization:

A polynomial model function of ten α_i parameters:

$$f(x) = \alpha_0 + \alpha_1 x + \alpha_2 x^2 + \cdots + \alpha_{10} x^{10}$$

Calculate the OLS estimate, and the OLS ridge-regularized estimates for the parameters given the sample points of the graph of f given that the values are y = (-82, -822755942135606300), (93, 2430665097409248000), (-23, -9261948480746.87), (37, 644396936880764.5), (-34, -308214284425589.6), (63, 79132774929002450), (94, 2701442110691023400), (-69, -183005635711177540), (-54, -20767364481378264), (62, 67147394829612940), (-20, -2882691880287.22), (-77, -475543956055209100), (15, 165941225818.79), (85, 1167049312831448600), (-76, -413829240205171600), (-37, -648874738043371.5), (89, 1651784866240920600), (-77, -493275386555451600), (-18, -1080503519688.94), (-46, -4822540268120077).

Remember to include the steps of your computation which are more important than the actual computations.

The objective of this task is to calculate the OLS (Ordinary Least Squares) estimate and the OLS ridge-regularized estimates for the parameters of a polynomial model function of ten αi parameters using a given set of sample points of an unknown function f.

To define a model function $f(x)$ as a polynomial function of ten αi parameters, the function should take a value x as input and return the value of $f(x)$ as output.

The given polynomial function is:

$$f(x) = \alpha_0 + \alpha_1 x + \alpha_2 x^2 + \cdots + \alpha_{10} x^{10} \ldots\ldots (5.1)$$

The given sample points are:

x	y
-82	-822755942135606300
93	2430665097409248000
-23	-9261948480746.87
37	644396936880764.5
-34	-308214284425589.6

x	y
-20	-2882691880287.22
-77	-475543956055209100
15	165941225818.79
85	1167049312831448600
-76	-413829240205171600

63	79132774929002450	-37	-648874738043371.5
94	2701442110691023400	89	1651784866240920600
-69	-183005635711177540	-77	-493275386555451600
-54	-20767364481378264	-18	-1080503519688.94
62	67147394829612940	-46	-4822540268120077

Table 5.2

Referring to section 6.2 of the coursebook, to demonstrate how the OLS method estimates the parameters of the assumed model, we will explore this in the specific case of linear regression. Since the random variable x is a vector, we are dealing with multiple linear regression.

Let's calculate the OLS estimate for the parameters by finding the least squares solution of equation *(5.1)* using linear algebra. This is the solution that minimizes the sum of the squared differences between the predicted values and the actual values.

Using the given set of data from *Table 5.2* and the model equation *(5.1)*, we can find the OLS estimate as follows.

Let's calculate the various powers for each sample point of x from x^0 to x^{20}. Next, we obtain the summation for various powers of x from x^0 to x^{20}. Now, we can create a 11×11 matrix A with each row and column containing the summation of the sample x values starting from x^0 to x^{20}.

Next, let's create a matrix B that contains the corresponding output values for each sample point. Each output value is obtained by the summation of sample y points multiplied by the summation of various powers of x.

To obtain the values for $\alpha_0, \alpha_1 \ldots \ldots \alpha_{10}$ we solve the simultaneous equations by first computing the multiplicative inverse of matrix A as A^{-1} and then calculating the dot product of the matrix A^{-1} with matrix B $(\mathrm{Stingher, 2019})$:

The simultaneous equations are as follows:

$$\sum y = \alpha_0 + \alpha_1 \sum \mathrm{x} + \alpha_2 \sum \mathrm{x}^2 + \alpha_3 \sum x^3 + \alpha_4 \sum x^4 + \alpha_5 \sum \mathrm{x}^5 + \alpha_6 \sum \mathrm{x}^6 + \alpha_7 \sum \mathrm{x}^7 + \alpha_8 \sum \mathrm{x}^8 + \alpha_9 \sum \mathrm{x}^9 + \alpha_{10} \sum \mathrm{x}^{10}$$

$$\sum xy = \alpha_0 \sum \mathrm{x} + \alpha_1 \sum \mathrm{x}^2 + \alpha_2 \sum x^3 + \alpha_3 \sum x^4 + \alpha_4 \sum \mathrm{x}^5 + \alpha_5 \sum \mathrm{x}^6 + \alpha_6 \sum \mathrm{x}^7 + \alpha_7 \sum \mathrm{x}^8 + \alpha_8 \sum \mathrm{x}^9 + \alpha_9 \sum \mathrm{x}^{10} + \alpha_{10} \sum \mathrm{x}^{11}$$

.

.

.

$$\sum x^{10}\,\mathrm{y} = \alpha_0 \sum \mathrm{x}^{10} + \alpha_1 \sum \mathrm{x}^{11} + \alpha_2 \sum \mathrm{x}^{12} + \alpha_3 \sum \mathrm{x}^{13} + \alpha_4 \sum \mathrm{x}^{14} + \alpha_5 \sum \mathrm{x}^{15} + \alpha_6 \sum \mathrm{x}^{16} + \alpha_7 \sum \mathrm{x}^{17} + \alpha_8 \sum \mathrm{x}^{18} + \alpha_9 \sum \mathrm{x}^{19} + \alpha_{10} \sum \mathrm{x}^{20}$$

The various α values are as follows:

$\alpha_0 = 1.03761988\mathrm{e} + 15;$ $\alpha_1 = 8.63838904\mathrm{e} + 14;$ $\alpha_2 = 5.63209149\mathrm{e} + 13;$

$\alpha_3 = -9.95403256e + 11;$ $\alpha_4 = -7.92971021e + 10;$ $\alpha_5 = 1.63694031e + 08;$

$\alpha_6 = 3.15537850e + 07;$ $\alpha_7 = 2.73057141e + 04;$ $\alpha_8 = -4.78941707e + 03;$

$\alpha_9 = 2.37938795e - 01;$ $\alpha_{10} = 2.49667964e - 01;$

Referring to section 6.2 of the coursebook, the OLS ridge-regularized estimates for the parameters can be found by minimizing the joint cost function of the model and the penalty, given by:

$$\hat{a} = \arg\min[\sum_{i=1}^{N}(y_i - \hat{y_i})^2 + \text{penalty}(a)]$$

Where, $\arg\min$ indicates the least squares method formulated as a optimization problem.

$$\text{penalty}(a) = \lambda \sum_{k=1}^{K} a_k^2$$

Where, λ is a free parameter that determines the strength of the regularization.

Finally, let's use the calculated α values in the polynomial model function *(5.1)* to obtain the predictions.

The function takes the value of x as input and returns $f(x)$ as output as follows:

$f(-82) = -8.209553022061e + 17;$ $f(93) = 2.439382658479653e + 18;$

$f(-23) = 4042961014082352.5;$ $f(37) = -9622207924744922.0;$

$f(-34) = 1750613617121832.8;$ $f(63) = 8.490509371346051e + 16;$

$f(94) = 2.6911180492098867e + 18;$ $f(-69) = -1.734793836284388e + 17;$

$f(-54) = -2.28301044616286e + 16;$ $f(62) = 6.757357094093685e + 16;$

$f(-20) = 2905393228609914.5;$ $f(-77) = -4.812963184930025e + 17;$

$f(15) = 1.976977380338506e + 16;$ $f(85) = 1.1526208000605266e + 18;$

$f(-76) = -4.28233615908522e + 17;$ $f(-37) = -654559529202721.0;$

$f(89) = 1.6649913926982687e + 18;$ $f(-77) = -4.812963184930025e + 17;$

$f(-18) = 1912646601006984.0;$ $f(-46) = -9994621939405044.0;$

A custom function is written in python on jupyter notebook to perform the following computations:

- $\text{array}()$ function from the numpy library is used to transform to given sample $\mathbf{x}$ and y values into a numpy array.
- $\text{power}()$ function from numpy library is used to calculate the various powers of sample x values.
- Simultaneous equations are computed using linear algebra functions in the numpy library. $\text{Inv}()$ function is used to calculate the multiplicative inverse of matrix A as A^{-1}. $\text{dot}()$ function is used to calculate the dot product of matrix A^{-1} and matrix B
- Matplotlib library is used to plot
  - Sample values of x and y
  - Sample value of x and the predictions $\hat{y}$
  - Line graph to that compares y and $\hat{y}$

Here is the custom function that performs the above-mentioned computations and visualizations:

```
import numpy as np
import pandas as pd
import plotly.express as px
import matplotlib.pyplot as plt

x_arr=np.array([-82,93,-23,37,-34,63,94,-69,-54,62,-20,-77,15,85,-76,-
37,89,-77,-18,-46])

#Calculation of various powers for x
x1 = np.float_(x_arr);x0=np.power(x1,0);x2=np.power(x1, 2);x3=np.power(x1,
3);x4=np.power(x1, 4);x5=np.power(x1, 5);
x6=np.power(x1, 6);x7=np.power(x1, 7);x8=np.power(x1, 8);x9=np.power(x1,
9);x10=np.power(x1, 10);x11=np.power(x1, 11);
x12=np.power(x1, 12);x13=np.power(x1, 13);x14=np.power(x1,
14);x15=np.power(x1, 15);x16=np.power(x1, 16);
x17=np.power(x1, 17);x18=np.power(x1, 18);x19=np.power(x1,
19);x20=np.power(x1, 20);

#Calculation of summation x
sx1=sum(x1);sx0=sum(x0);sx2=sum(x2);sx3=sum(x3);sx4=sum(x4);sx5=sum(x5);sx6
=sum(x6);sx7=sum(x7);sx8=sum(x8);
sx9=sum(x9);sx10=sum(x10);sx11=sum(x11);sx12=sum(x12);sx13=sum(x13);sx14=su
m(x14);sx15=sum(x15);sx16=sum(x16);sx17=sum(x17);
sx18=sum(x18);sx19=sum(x19);sx20=sum(x20);

y_arr=np.array([-822755942135606300,2430665097409248000,-
9261948480746.87,644396936880764.5,-308214284425589.6,79132774929002450,
   2701442110691023400,-183005635711177540,-
20767364481378264,67147394829612940,-2882691880287.22,-475543956055209100,
   165941225818.79,1167049312831448600,-413829240205171600,-
648874738043371.5,1651784866240920600,-493275386555451600,
   -1080503519688.94,-4822540268120077])
y = np.float_(y_arr)

#Multiplication of x powers and y values
x1y=x1*y;x2y=x2*y;x3y=x3*y;x4y=x4*y;x5y=x5*y;x6y=x6*y;x7y=x7*y;x8y=x8*y;x9y
=x9*y;x10y=x10*y;
xylst=[x1y,x2y,x3y,x4y,x5y,x6y,x7y,x8y,x9y,x10y]

#Calculation of summation y
sy=sum(y);sx1y=sum(x1y);sx2y=sum(x2y);sx3y=sum(x3y);sx4y=sum(x4y);sx5y=sum(
x5y);sx6y=sum(x6y);sx7y=sum(x7y);
sx8y=sum(x8y);sx9y=sum(x9y);sx10y=sum(x10y);

#Matrix A
A=[[1,sx1,sx2,sx3,sx4,sx5,sx6,sx7,sx8,sx9,sx10],[sx1,sx2,sx3,sx4,sx5,sx6,sx
7,sx8,sx9,sx10,sx11],

[sx2,sx3,sx4,sx5,sx6,sx7,sx8,sx9,sx10,sx11,sx12],[sx3,sx4,sx5,sx6,sx7,sx8,s
x9,sx10,sx11,sx12,sx13],

[sx4,sx5,sx6,sx7,sx8,sx9,sx10,sx11,sx12,sx13,sx14],[sx5,sx6,sx7,sx8,sx9,sx1
0,sx11,sx12,sx13,sx14,sx15],
```

```
[sx6,sx7,sx8,sx9,sx10,sx11,sx12,sx13,sx14,sx15,sx16],[sx7,sx8,sx9,sx10,sx11
,sx12,sx13,sx14,sx15,sx16,sx17],

[sx8,sx9,sx10,sx11,sx12,sx13,sx14,sx15,sx16,sx17,sx18],[sx9,sx10,sx11,sx12,
sx13,sx14,sx15,sx16,sx17,sx18,sx19],
  [sx10,sx11,sx12,sx13,sx14,sx15,sx16,sx17,sx18,sx19,sx20]]

#Matrix B
B=np.array([sy,sx1y,sx2y,sx3y,sx4y,sx5y,sx6y,sx7y,sx8y,sx9y,sx10y])

#Calculation of alpha values
alp = np.linalg.inv(A).dot(B)

#Calculation of ycap values
ycap=[]
for rec in range(0,20):

listval=alp[0]+alp[1]*x1[rec]+alp[2]*x2[rec]+alp[3]*x3[rec]+alp[4]*x4[rec]+
alp[5]*x5[rec]+alp[6]*x6[rec]+alp[7]*x7[rec]+alp[8]*x8[rec]+alp[9]*x9[rec]+
alp[10]*x10[rec]
    ycap.append(listval)

#x and y plot
fig, ax = plt.subplots()
ax.plot(x_arr, y_arr, 'o');plt.title("x and y
plot");plt.xlabel("x");plt.ylabel("y")
plt.show()

#x and ycap plot
fig, ax = plt.subplots()
ax.plot(x_arr, ycap, 'o');plt.title("x and ycap
plot");plt.xlabel("x");plt.ylabel("ycap")
plt.show()

# comparison of y and ycap
df = pd.DataFrame(data=ycap, index=y_arr)
compplt = px.line(df, x=y_arr, y=ycap, title='y and ycap comparison',
                labels=dict(x="original", y="estimate")
                 )
compplt.show()
```

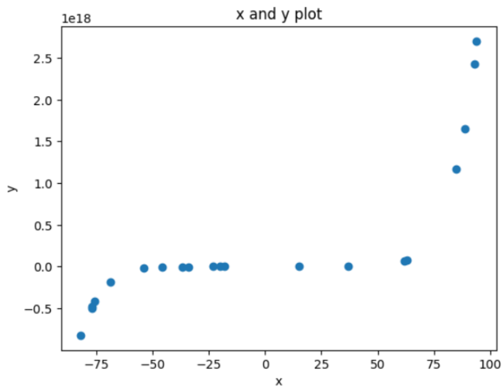

Figure 4: x and y plot generated in jupyter-notebook

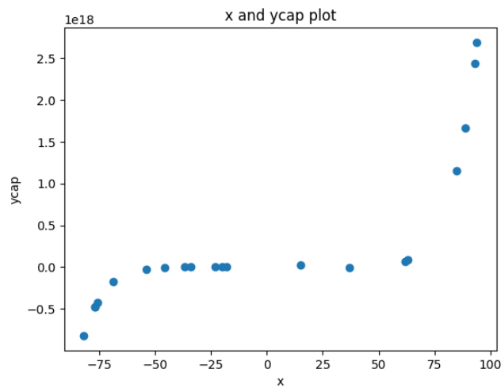

Figure 5: x and ycap plot generated in jupyter-notebook

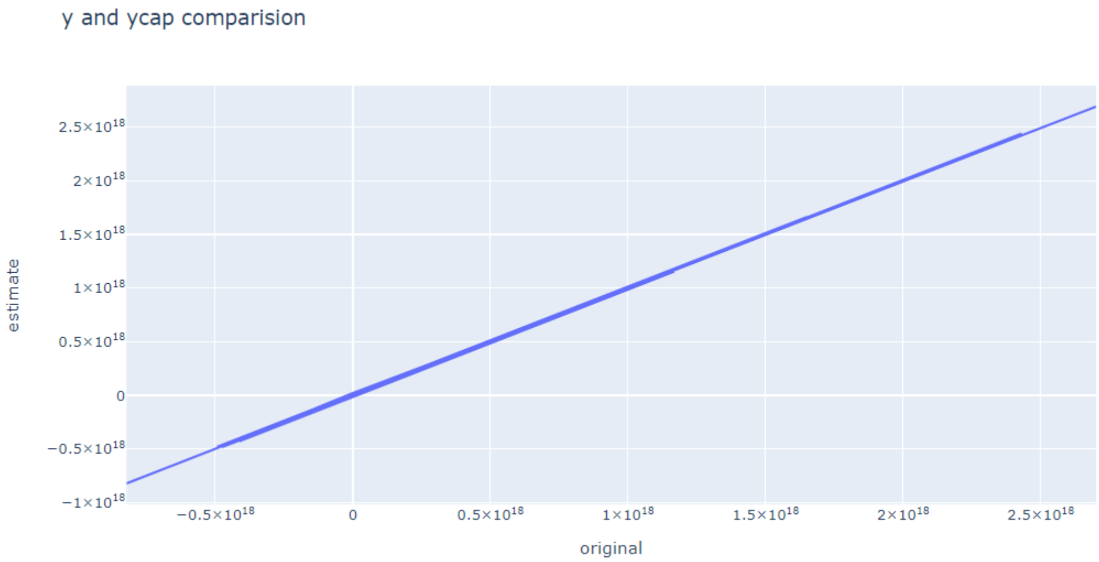

Figure 6: y and ycap comparison generated in jupyter-notebook

Task 6: Bayesian Estimates

(following Hogg, McKean & Craig, exercise 11.2.2)

Let $X_1, X_2, \dots, X_{10}$ be a random sample from a gamma distribution with $\alpha = 3$ and $\beta = 1/\theta$. Suppose we believe that θ follows a gamma-distribution with $\alpha = 42$ and $\beta = 62$ and suppose we have a trial $(x_1, \dots, x_n)$ with an observed $\bar{x} = 31.94$

a) Find the posterior distribution of θ.

b) What is the Bayes point estimate of θ associated with the square-error loss function?

c) What is the Bayes point estimate of θ using the mode of the posterior distribution?

The probability density function(PDF) for a Gamma distribution $X \sim \Gamma(\alpha, \beta)$ is given by(“Gamma Distribution,” 2022):

$$f_X(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} \qquad (6.1)$$

From the equation *(6.1)*, the gamma distribution for a random sample with $\alpha = 3$ and $\beta = 1/\theta$ is given by:

$$f_X(x) = \frac{1}{\theta^3\Gamma(3)} x^2 e^{-\frac{x}{\theta}}$$

Since we believe that θ follows a gamma-distribution which is the same as the original distribution, the prior probability distribution(“Prior Probability,” 2023) is transformed into a posterior distribution(“Posterior Probability,” 2023) with $\alpha = 42$ and $\beta = 62$ derived from the equation *(6.1)* as follows:

$$f_X(x) = \frac{x^{\alpha-1}e^{-\frac{x}{\theta}}}{\Gamma(\alpha)\theta^{\alpha}} \Longrightarrow \frac{x^{41}e^{-62x}}{\Gamma(42)62^{42}} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \textit{(6.2)}$$

As per section 6.1 of the coursebook, the likelihood is the probability of observing a given dataset or the joint distribution of the dataset evaluated at the given data. Therefore, the Bayesian likelihood function θ for the 10 random samples $X_1, X_2, \ldots, X_{10}$ with $\alpha = 3$ and $\beta = 1/\theta$ referring to the equation *(6.2)* is given by:

$$L(\theta) = \frac{x_1^2 e^{-\frac{x_1}{\theta}}}{\Gamma(3)\theta^3} \cdot \frac{x_2^2 e^{-\frac{x_2}{\theta}}}{\Gamma(3)\theta^3} \ldots \ldots \frac{x_{10}^2 e^{-\frac{x_{10}}{\theta}}}{\Gamma(3)\theta^3}$$

$$L(\theta) = \frac{\sum_{i=1}^{10}(x_i)^2 e^{-\frac{x_i}{\theta}}}{\Gamma(3)\theta^3} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \textit{(6.3)}$$

Since we know that the observed mean value $\bar{x} = 31.94$ and there are 10 random samples, the value of $X_1, X_2, \ldots, X_{10}$ is given by(“Expected Value,” 2022b):

$$X_1, X_2, \ldots, X_{10} = \bar{x} \times 10 \Rightarrow 31.94 \times 10 = 319.4 \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \textit{(6.4)}$$

Substituting the value of $X_1, X_2, \ldots, X_{10}$ from the above equation in equation *(6.3)* we get:

$$L(\theta) = \frac{(319.4)^2 e^{-\frac{(319.4)}{\theta}}}{\Gamma(3)\theta^3} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \textit{(6.5)}$$

As per section 6.1 of the coursebook, a value θ that maximizes the likelihood of observing the data is called the maximum likelihood estimate. Therefore, to obtain the maximum likelihood for the given task, we will be finding a parameter θ that maximizes the probability of observing the given data by adjusting the value of the parameter(s) θ for an assumed choice of a probability distribution.

Since the logarithm of a product is the sum of the logarithms and it makes our likelihood function easier to handle in practice, let’s take log on both sides for the equation *(6.5)* as follows:

$$\log\,[L(\theta)] = \log\left[\frac{(319.4)^2 e^{-\frac{(319.4)}{\theta}}}{\Gamma(3)\theta^3}\right] \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \textit{(6.6)}$$

$$\log[L(\theta)] = 2\log(319.4) - \frac{(319.4)}{\theta} - \log\Gamma(3) - 3\log\theta \ldots\ldots\ldots\ldots\ldots\ldots \textit{(6.7)}$$

As per section 6.1 of the coursebook, we can compute the first derivative for the equation *(6.7)* to obtain the maximum likelihood estimate given by:

$$\frac{d}{d\theta}L(\theta) = \frac{d}{d\theta}\left[2\log(319.4) - \frac{(319.4)}{\theta} - \log\Gamma(3) - 3\log\theta\right] = 0$$

$$\frac{d}{d\theta}L(\theta) = 0 + \frac{319.4}{\theta^2} - 0 - \frac{3}{\theta} = 0 \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\textit{(6.8)}$$

$$\theta = \frac{319.4}{3} = 106.46\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\textit{(6.9)}$$

Equation *(6.9)* is the Bayes point estimate of θ associated with the square-error loss function.

The Bayes point estimate of θ using the mode of the posterior distribution is the maximizer of the likelihood function which is obtained by differentiating equation *(6.8)* as follows:

$$\frac{d^2}{d\theta^2}L(\theta) = \frac{d}{d\theta}\left[\frac{319.4}{\theta^2} - \frac{3}{\theta}\right]$$

$$\frac{d^2}{d\theta^2}L(\theta) = -\frac{638.8}{\theta^3} + \frac{3}{\theta^2}$$

Substituting the value of θ from the equation *(6.9)* in the above equation we get:

$$\frac{d^2}{d\theta^2}L(\theta) = -\frac{638.8}{106.46^3} + \frac{3}{106.46^2}$$

$$\frac{d^2}{d\theta^2}L(\theta) = -2.64 \times 10^{-4} < 0\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\textit{(6.10)}$$

Referring to section 6.1 of the coursebook, the equation *(6.10)* is indeed a maximizer of the likelihood function because the second derivative is negative here.

References

Alternative hypothesis. (2022). In *Wikipedia*. https://en.wikipedia.org/w/index.php?title=Alternative_hypothesis&oldid=1122934976 Last accessed: 2022-11-20

Central limit theorem. (2022). In *Wikipedia*. https://en.wikipedia.org/w/index.php?title=Central_limit_theorem&oldid=1129090135 Last accessed: 2023-01-08

Confidence interval. (2022). In *Wikipedia*. https://en.wikipedia.org/w/index.php?title=Confidence_interval&oldid=1128642915 Last accessed: 2023-01-09

Cumulative distribution function. (2023). In *Wikipedia*. https://en.wikipedia.org/w/index.php?title=Cumulative_distribution_function&oldid=1136509526 Last accessed: 2023-01-31

Expected value. (2022a). In *Wikipedia*. https://en.wikipedia.org/w/index.php?title=Expected_value&oldid=1124470444 Last accessed: 2023-01-06

Expected value. (2022b). In *Wikipedia*. https://en.wikipedia.org/w/index.php?title=Expected_value&oldid=1124470444 Last accessed: 2023-01-06

Gamma distribution. (2022). In *Wikipedia*. https://en.wikipedia.org/w/index.php?title=Gamma_distribution&oldid=1128893625 Last accessed: 2023-01-11

HelpYourMath - Statistics (Director). (2019, November 26). *Lesson 13.5 z- test left tail test*. https://www.youtube.com/watch?v=VXae5k_4xsI Last accessed: 2023-01-09

Hogg, R. V., McKean, J. W., & Craig, A. T. (2005). *Introduction to mathematical statistics* (6th ed). Pearson Education.

Normal distribution. (2023). In *Wikipedia*. https://en.wikipedia.org/w/index.php?title=Normal_distribution&oldid=1131761789 Last accessed: 2023-01-08

Null hypothesis. (2023). In *Wikipedia*. https://en.wikipedia.org/w/index.php?title=Null_hypothesis&oldid=1132319250 Last accessed: 2023-01-09

Poisson distribution. (2023). In *Wikipedia*. https://en.wikipedia.org/w/index.php?title=Poisson_distribution&oldid=1136434953 Last accessed: 2023-01-30

Posterior probability. (2023). In *Wikipedia*. https://en.wikipedia.org/w/index.php?title=Posterior_probability&oldid=1131306782 Last accessed: 2023-01-03

Prior probability. (2023). In *Wikipedia*. https://en.wikipedia.org/w/index.php?title=Prior_probability&oldid=1132804087 Last accessed: 2023-01-10

Probability density function. (2022). In *Wikipedia*. https://en.wikipedia.org/w/index.php?title=Probability_density_function&oldid=1130726352 Last accessed: 2023-01-06

Probability mass function. (2022). In *Wikipedia*. https://en.wikipedia.org/w/index.php?title=Probability_mass_function&oldid=1124446867 Last accessed: 2023-01-05

Quartile. (2022). In *Wikipedia*. https://en.wikipedia.org/w/index.php?title=Quartile&oldid=1128289583 Last accessed: 2023-01-05

Statistical significance. (2022). In *Wikipedia*. https://en.wikipedia.org/w/index.php?title=Statistical_significance&oldid=1129611776 Last accessed: 2023-01-09

Stingher, C. (2019, November 15). *Answer to “Solving simultaneous equations with Python.”* Stack Overflow. https://stackoverflow.com/a/58878679 Last accessed: 2023-01-11

The Organic Chemistry Tutor (Director). (2019, October 3). *One Tailed and Two Tailed Tests, Critical Values, & Significance Level—Inferential Statistics*. https://www.youtube.com/watch?v=XHPIEp-3yC0 Last accessed: 2023-01-09

Variance. (2022). In *Wikipedia*. https://en.wikipedia.org/w/index.php?title=Variance&oldid=1117946674 Last accessed: 2023-01-05